

**Economics Research Group**  
**IBMEC Business School – Rio de Janeiro**  
**<http://professores.ibmecrj.br/erg>**

**IBMEC RJ ECONOMICS DISCUSSION PAPER 2011-02**

*Full series available at <http://professores.ibmecrj.br/erg/dp/dp.htm>.*

# **BAYESIAN FACTOR SELECTION IN DYNAMIC TERM STRUCTURE MODELS**

**Márcio Poletti Laurini**

# BAYESIAN FACTOR SELECTION IN DYNAMIC TERM STRUCTURE MODELS

MÁRCIO POLETTI LAURINI  
IBMEC RJ

**ABSTRACT.** This paper discusses Bayesian procedures for factor selection in dynamic term structure models through simulation methods based on Markov Chain Monte Carlo. The number of factors, besides influencing the fitting and prediction of observed yields, is also relevant to features such as the imposition of no-arbitrage conditions. We present a methodology for selecting the best specification in the Nelson-Siegel class of models using Reversible Jump MCMC.

**Keywords:** Dynamic Term Structure Models, Model Selection, Reversible Jump MCMC.

## 1. INTRODUCTION AND METHODOLOGY

Selecting the appropriate specification in econometric models is fundamental to guarantee the expected econometric properties, such as consistency, absence of bias and efficiency. The functional form and variables included in the specification chosen by the researcher also has important practical consequences, such as the predictive performance of the model, and in models used in financial applications, the estimated specification can affect the asset prices derived from the model.

In the analysis of dynamic econometric models for the term structure of interest rates, the functional form chosen, besides the influence on the properties of estimators and forecasts, may also have important consequences on the validity of no-arbitrage conditions. The functional form chosen is directly associated with the consistency with no-arbitrage in the sense defined in Filipovic (1999) and Filipovic (2001). For example in the widely used family of Nelson-Siegel (Nelson & Siegel (1987)) models is possible to show that original Nelson-Siegel specification is not consistent with no-arbitrage, but the addition of an additional slope factor, leading to the model proposed by Svensson (1994), allows a specification consistent with no-arbitrage (although limited in the empirical fit), and the overall consistency is achieved by an additional latent factor, leading to the five factor model of Björk & Christensen (1999).

These conditions were then reformulated in the affine formulation for the Nelson-Siegel family proposed in Christensen *et al.* (2010), Christensen *et al.* (2009) and Joslin *et al.* (2011). In this formulation consistence with no-arbitrage is obtained when the model contains three or five factors, but not in other specifications. In this formulation the existence of no-arbitrage conditions is linked to the matching of each slope factor with a associated curvature factor. This way, the imposition of no-arbitrage conditions depends on the validity of appropriate specification, related to the number of factors chosen. It

is also important to note that the arbitrage free formulation of Christensen *et al.* (2010) and Christensen *et al.* (2009), consistence with the no-arbitrage is obtained by adding an correction factor in the Nelson-Siegel family. For example in the formulation of the model with five factors the arbitrage free curve is given by:

$$(1.1) \quad y(t, T) = X_t^1 + \frac{1 - e^{-\lambda_1(T-t)}}{\lambda_1(T-t)} X_t^2 + \frac{1 - e^{-\lambda_2(T-t)}}{\lambda_2(T-t)} X_t^3 + \left[ \frac{1 - e^{-\lambda_1(T-t)}}{\lambda_1(T-t)} - e^{-\lambda_1(T-t)} \right] X_t^4 + \left[ \frac{1 - e^{-\lambda_2(T-t)}}{\lambda_2(T-t)} - e^{-\lambda_2(T-t)} \right] X_t^5 - \frac{C(t, T)}{T-t}$$

with  $y(t, T)$  denoting the observed yields,  $(T-t)$  the time to maturity,  $X_t^1, X_t^2, X_t^3, X_t^4, X_t^5$  the five latent factors and  $\lambda_1, \lambda_2$  the decay factors of the Nelson-Siegel model. In this specification correction of for no-arbitrage is given by the term  $C(t, T)$ , which is basically a function<sup>1</sup> of the parameters  $\lambda_1$  and  $\lambda_2$ . Because of this term  $C(t, T)$  the procedure of model selection is crucial in the analysis of consistency with no-arbitrage. For example the models with 3 factors (including  $X_t^1, X_t^2$  and  $X_t^4$ ) is arbitrage-free, but the models with only one factor ( $X_t^1$ ) and four factors ( $X_t^1, X_t^2, X_t^4$  and  $X_t^5$ ) are not consistent with no-arbitrage.

Note that the usual procedures of model selection are hampered by the nonlinear formulation that contains common parameters in the definition of the latent factors. Each decay parameter  $\lambda_i$  affects both the associated slope and curvature factors. Besides the trivial case of misspecification raised by the omission of a significant latent factor, note that the factor analysis with a redundant (e.g. correct model factors contains four factors and the estimated model contains five factors), due to the nonlinearity of the model, the consistency of the estimation is no longer valid, as it would with a specification of a linear regression model. This problem hinders the use of traditional methods of model selection such as information criteria, which are only consistent (BIC by example, e.g. Claeskens & Hjort (2008)) in the selection of linear models, property which is not generally valid for nonlinear models. Another difficulty in using information criteria is that the validity these procedures is only asymptotic and finite sample properties may not be optimal.

Because of these difficulties, this paper proposes the use of Bayesian methods based on Reversible Jump Markov Chain Monte Carlo. This methodology generalizes the sampling framework of Markov Chain Monte Carlo (eg Gamerman & Lopes (2006), Robert & Casella (2005)) to perform the estimation of models whose dimension is variable. In this situation we can adequately address the problem of selection of variables and models, and perform procedures such as model averaging. A Bayesian model of variable dimension<sup>2</sup> can be defined as an indexed collection of models:

$$(1.2) \quad \mathcal{M}_k = \{f(\bullet|\theta_k); \theta_k \in \Theta_k\}$$

<sup>1</sup>The analytical form of this term is quite extensive, and therefore omitted from the article, but can be found in Christensen *et al.* (2010) and Christensen *et al.* (2009).

<sup>2</sup>We follow the definition adopted in Robert & Casella (2005).

associated with a collection of priors on the parameters of these models, denoted by  $\pi_k(\theta_k)$ . It is also necessary define a prior on the index of the model -  $\varrho(k), k = 1, \dots, K$ . We join the notation for the structure of priors as  $\pi_k(k, \theta_k) = \varrho(k)\pi_k(\theta_k)$ . This is a proper density defined with respect to Lebesgue measure on the union of the spaces  $\Theta = \cup_k \{k\} \times \Theta_k$ . With this structure we can obtain the posterior distribution of parameters via Bayes' lemma:

$$(1.3) \quad p(\mathcal{M}_i|x) = \frac{\varrho(i) \int_{\Theta_i} f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i}{\sum_j \varrho(j) \int_{\Theta_j} f_j(x|\theta_j)\pi_j(\theta_j)d\theta_j}.$$

From Bayes' lemma we can choose the model with the highest  $p(\mathcal{M}_i|x)$ , or then perform a procedure of model averaging:

$$(1.4) \quad \sum_j p(\mathcal{M}_i|x) \int_{\Theta_j} f_j(x|\theta_j)\pi_j(\theta_j)d\theta_j$$

with  $x$  denoting the observed sample and the above equation as a predictive density. Although this structure is a Bayesian complete model, there are several difficulties in implementing this computational problem, since such methods as Gibbs sampling can not be directly used in this context, since it is necessary to define how they will be made the transitions between models  $\mathcal{M}_i$  considered in the sampling procedure<sup>3</sup>.

The methodology known as Reversible Jump MCMC, proposed by Green (1995) is based on the determination of a reversible transition kernel between models, i.e., a kernel satisfying  $\int_A \int_B K(x, dy)\pi(x)dx = \int_A \int_B K(y, dx)\pi(y)dy$ , for  $A, B \subset \theta$  and  $\pi$  representing the invariant density of the model. The kernel represents the transition between different models using the following decomposition:

$$(1.5) \quad K(x, B) = \sum_m \int_B \rho_m(x, y')q_m(x, dy') + \omega(x)\mathbb{I}_B(x)$$

with  $q_m$  denoting the transition function for the model  $\mathcal{M}_m$ , with transition probabilities  $\rho_m$ . We define the probability of not transition as  $\omega(x) = 1 - \sum_m (q_m, p_m)(x, \Theta_m)$ . Under some technical conditions we can define the probability of transition inside the usual Metropolis-Hastings procedure:

$$(1.6) \quad \rho_m(x, y) = \min \left\{ 1, \frac{g_m(y, x)}{g_m(x, y)} \right\}$$

with  $g_m(\cdot)$  denoting the instrumental density.

The algorithm proposed by Green (1995) solves the computational problem of the change (jump) between the different models considered through supplementation of the spaces  $\Theta_{k_1}$  and  $\Theta_{k_2}$  with artificial spaces, creating a bijection between the models. For example this solves the problem of distinct parameter's dimension between each model

<sup>3</sup>See Robert & Casella (2005) for a discussion of problems involved

considered, and also allows to create the necessary Jacobians of these transformations, by the use of instrumental models with a fixed dimension, usually taken by the size of the largest model considered. The details of computational implementation used in this work can be found in Lunn *et al.* (2009), which discusses the necessary changes in Gibbs and Metropolis-Hastings sampling procedures to implement the Reversible Jump MCMC of Green (1995).

To implement the Reversible Jump MCMC procedure for model selection in the family of Nelson-Siegel models, we perform some modifications on the Reversible Jump MCMC procedure proposed in Green (1995) and Lunn *et al.* (2009), using some special features of this family of models. Note that conditional to the parameters of decay  $\lambda_i$ , the specification given by equation 1.1 is a linear model. The second approach is the ability to analyze the average behavior of the relevant factors, thus avoiding direct estimation of latent factors, as in Diebold & Li (2006), Christensen *et al.* (2010) and Christensen *et al.* (2009). This simplification allows performing model selection through the methodology of Reversible Jump MCMC developed for selection of linear models.

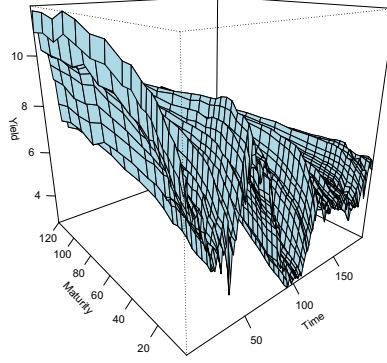
The procedure is performed through the use of the methodology of Gibbs Sampling and Metropolis-Hastings proposed in Lunn *et al.* (2009), with the modification for the estimation of a conditional linear model. In first stage of the estimation of each sub-model, we estimate the parameters  $\lambda_i$  for each proposed model by Metropolis-Hastings sampling. Conditional to the estimation these parameters, the model becomes conditionally linear and we can perform the Reversible Jump algorithm to the estimation of loadings  $\bar{X}^1, \bar{X}^2, \bar{X}^3, \bar{X}^4$  and  $\bar{X}^5$ . The selection of these loadings is performed assuming that these parameters are constant throughout the sample analyzed, and so we analyze the influence of each factor by average effect. Note that this assumption is not restrictive, being based only on the assumption of existence of an invariant distribution for these factors.

Another exploited property is to assume that the  $\bar{X}^1$  is present in all specifications. Because this factor has a level interpretation in the yield curve, it is always present in all specifications, and thus the model selection procedure may be restricted to the combinations between factors  $\bar{X}^2, \bar{X}^3, \bar{X}^4, \bar{X}^5$ , reducing the space of models analyzed. Another important point is that given the interpretation of slope and curvature of the other latent factors, they are approximately orthogonal, which avoids the transformation of regressors to orthogonal coordinates, as discussed in Robert & Casella (2005).

## 2. DATABASE AND RESULTS

This study used the same database of Diebold & Li (2006), Christensen *et al.* (2010) and Christensen *et al.* (2009). The base is composed of unsmoothed forward rates of U.S. Treasuries obtained by procedure of Fama & Bliss (1987), with fixed maturities of 1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, using monthly observations for the period 1985:01 - 2000:12, containing 192 observations of this curve. Figure 2.1 shows the database used. Descriptive statistics and other properties can be found in Diebold & Li (2006).

FIGURE 2.1. Fama-Bliss Database

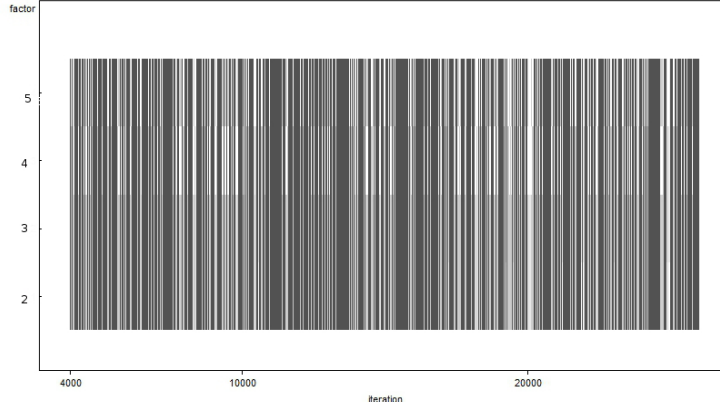


To complete the Bayesian specification, we define the set of priors for the parameters of interest. We assume that the priors for the decay parameters are  $\lambda_1 \sim \logNormal(-2.337901, 1)$ ,  $\lambda_2 \sim \logNormal(-0.1768565, 1)$ , consistent with the values estimated in Diebold & Li (2006), Christensen *et al.* (2010) and Christensen *et al.* (2009). For the factors  $\bar{X}^2, \bar{X}^3, \bar{X}^4, \bar{X}^5$  we assume a multivariate gamma-normal distribution  $p(\tau, \bar{X}|k) = Ga(\tau|, a, b) \times MVN(\bar{X}|\mu, \tau^{-1}I_k)$  where  $\mu = 0$  and  $I_k$  is an identity matrix of dimension  $k$ , with the dimension of the sub-model being evaluated, which is equivalent to imposing a same precision for all elements of the normal distribution multivariate analysis. In this problem assume a precision of  $\tau^{-1} = .0001$ , consistent with is observed for this data set. The hyperparameters  $a$  and  $b$  of Gamma prior are given by the .75 and .75. The last relevant prior is that defines the number of factors included in each specification. For the reported results, we use  $\varrho(k) \sim Binomial(.8, 4)$ . Other specifications for these priors were tested and the results do not change in a relevant way.

To obtain the posterior distributions, we run the MCMC procedure using 50,000 initial iterations (burn-in), and over 500,000 simulations to construct the posterior distributions. To illustrate the jumps between models of behavior, we put in Figure 2.2 a greyscale picture that shows which factors are selected in each iteration of the procedure. The results in the figure indicate the Reversible Jump MCMC procedure supports a specification with more factors, as can be noted by intensity (darker regions indicated that the factor is repeatedly selected) in this figure.

The posterior probabilities of each analyzed model are placed in Table 1. The column model structure shows the factors included in each specification. One denotes that the latent factor  $X^i$  is included, and zero that the factor ins not included in the specification - for example in the first line we have the posterior probability of the model with level, slope and curvature factors, according to the ordinance adopted in equation 1.1, corresponding

FIGURE 2.2. Jump Process

TABLE 1. Posterior Model Probabilities  $p(\mathcal{M}_i|x)$ 

model structure	posterior prob.	cumulative prob.
11010	9.0901E-5	0.00009
11011	9.5454E-4	0.00104
10110	0.001909	0.01622
10111	0.013272	0.002954
11100	0.032681	0.048909
11101	0.139181	0.188090
11110	0.156227	0.344318
11111	0.655681	1.000000

to the original Nelson-Siegel model. In the last column we find cumulative probability of all models analyzed.

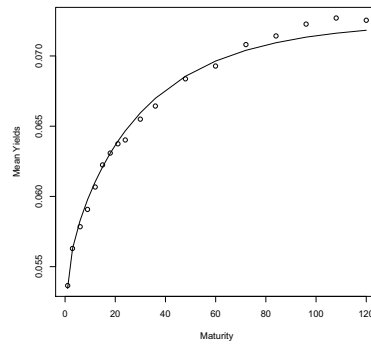
The results support a complete specification for this dataset, indicating that the model with five factors obtains the highest posterior probability in all specifications examined. Another way to analyze this result is through the marginal probabilities of each model, shown in Table 2. We obtain high marginal probabilities for all latent factors in the Nelson-Siegel family, supporting the use of models with five factors proposed in Björk & Christensen (1999) and Christensen *et al.* (2009) Christensen *et al.* (2009) and consistent with no arbitrage, an important financial result. Also we can see evidence in favor of a greater number of factors in the posterior distribution of parameter  $k$ , which indicates the number of factors included in each model. The posterior distribution of parameter  $k$  has posterior mean of 3.67, supporting the largest number of factors.

To illustrate the application of Reversible Jump MCMC methodology as a Bayesian Model Averaging procedure, we illustrate the model fit that would be obtained by weighting all models analyzed by the estimated posterior probabilities, according to equation 1.4. Figure 2.3 show the average yields over time (marked with circles in figure), compared to

TABLE 2. Marginal Probabilities

Factor	marginal prob.
2	0.9850
3	0.9957
4	0.8213
5	0.8209

FIGURE 2.3. Model Fit - Bayesian Model Averaging



the fit obtained by weighting all models, in the continuous line. The results shows that weighted model provides an excellent fit for the average yields observed in this sample.

### 3. CONCLUSIONS

In this paper we introduce the use of Reversible Jump Markov Chain Monte Carlo methods for model and factor selection in dynamic term structure models. From the special structure of Nelson-Siegel family, it is possible to adapt procedures used in the selection of linear models to exploit all possible specifications derived from this family of models.

These specifications have important consequences on fitting, prediction and consistency with no-arbitrage, and the Reversible Jump MCMC procedure allows overcome the limitations in the usual procedures for model selection, such as only asymptotic consistency for linear models.

### REFERENCES

- Björk, T., & Christensen, B. J. 1999. Interest Rate Dynamics and Consistent Forward Rate Dynamics. *Mathematical Finance*, **9**, 323–348.
- Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. 2009. An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model. *Econometrics Journal*, **12**, 33–64.
- Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. 2010. The Affine Arbitrage free Class of Nelson-Siegel Term Structure Models. *Journal of Econometrics*, **forthcoming**.



- Claeskens, C., & Hjort, N. L. 2008. *Model Selection and Model Averaging*. Cambridge University Press.
- Diebold, F.X., & Li, C. 2006. Forecasting the Term Structure of Government Bond Yields. *Journal Of Econometrics*, **130**, 337–364.
- Fama, Eugene, & Bliss, Robert R. 1987. The information in long-maturity forward rates. *American Economic Review*, **77**(Sept.), 680–691.
- Filipovic, D. 1999. A Note on the Nelson-Siegel Family. *Mathematical Finance*, **9**(4), 349–359.
- Filipovic, D. 2001. *Consistency Problems for Heath-Jarrow-Morton Interest Rate Models*. Springer-Verlag.
- Gamerman, D., & Lopes, H. 2006. *Stochastic Simulation for Bayesian Inference*. Chapman and Hall.
- Green, P. 1995. Reversible Jump MCMC computation and Bayesian Model determination. *Biometrika*, **82**(4), 711–732.
- Joslin, S., Singleton, K.J., & Zhu, H. 2011. A New Perspective on Gaussian Dynamic Term Structure Models. *Review of Financial Studies*, *Forthcoming*.
- Lunn, David J., Best, Nicky, & Whittaker, John C. 2009. Generic reversible jump MCMC using graphical models. *Statistics and Computing*, **19**(December), 395–408.
- Nelson, C. R., & Siegel, A. F. 1987. Parsimonious Modelling of Yield Curves. *Journal of Business*, **60**(4), 473–489.
- Robert, C. P., & Casella, G. 2005. *Monte Carlo Statistical Methods*. Springer.
- Svensson, L. E. O. 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. *NBER Working Paper*.